

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/71961/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Attaf, Meriem, Huseby, Eric and Sewell, Andrew K. ORCID:
<https://orcid.org/0000-0003-3194-3135> 2015. β T cell receptors as predictors of health and disease. Cellular and Molecular Immunology 12 , pp. 391-399. 10.1038/cmi.2014.134 file

Publishers page: <http://dx.doi.org/10.1038/cmi.2014.134>
<<http://dx.doi.org/10.1038/cmi.2014.134>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



REVIEW

$\alpha\beta$ T cell receptors as predictors of health and disease

Meriem Attaf¹, Eric Huseby² and Andrew K Sewell¹

The diversity of antigen receptors and the specificity it underlies are the hallmarks of the cellular arm of the adaptive immune system. T and B lymphocytes are indeed truly unique in their ability to generate receptors capable of recognizing virtually any pathogen. It has been known for several decades that T lymphocytes recognize short peptides derived from degraded proteins presented by major histocompatibility complex (MHC) molecules at the cell surface. Interaction between peptide-MHC (pMHC) and the T cell receptor (TCR) is central to both thymic selection and peripheral antigen recognition. It is widely assumed that TCR diversity is required, or at least highly desirable, to provide sufficient immune coverage. However, a number of immune responses are associated with the selection of predictable, narrow, or skewed repertoires and public TCR chains. Here, we summarize the current knowledge on the formation of the TCR repertoire and its maintenance in health and disease. We also outline the various molecular mechanisms that govern the composition of the pre-selection, naive and antigen-specific TCR repertoires. Finally, we suggest that with the development of high-throughput sequencing, common TCR 'signatures' raised against specific antigens could provide important diagnostic biomarkers and surrogate predictors of disease onset, progression and outcome.

Cellular & Molecular Immunology advance online publication, 26 January 2015; doi:10.1038/cmi.2014.134

Keywords: T cell receptor (TCR); TCR repertoire; TCR diversity; TCR clonotype; TCR bias; Deep sequencing

INTRODUCTION

T cells orchestrate immune responses by interrogating protein expression *via* peptides cradled in major histocompatibility complex (MHC) molecules at the cell surface. The T cell receptor (TCR) is the fundamental unit underlying all peptide-MHC (pMHC) recognition events. In the thymus, T cell signalling induced by self-pMHC engagement contributes to the process of selection at the double-positive stage, whereby only a fraction of thymocytes bearing TCRs within a narrow affinity range are permitted to differentiate into mature T cells.¹ In secondary lymphoid organs, ligation of the TCR to pMHC provides the cell with the earliest signals required for the execution of a complex differentiation programme associated with effector function. In the steady state, homeostasis of T cell numbers is also MHC-dependent.²

The TCR is a heterodimer of one α and one β chain, or one γ and one δ chain, which are disulphide-linked.^{3,4} Each TCR chain is composed of a constant and a variable domain, followed by a membrane-spanning region and a short cytosolic tail. Diversity in the TCR is predominantly confined to six hypervariable hairpin loops in the variable domains, called complementarity-determining regions⁵ (CDR) (Figure 1a).

TCR chains are assembled somatically during T cell development by the joining of discrete V, (D) and J gene segments by recombination activating gene (RAG)1 and RAG2 (Figures 1b and 2). The process of V (D) J recombination is such that CDR1 α , CDR1 β , CDR2 α and CDR2 β are entirely encoded in germline DNA segments. In contrast, the CDR3 loops are the product of junctional diversity and are consequently hypervariable (Figure 1b).

Gene rearrangement is an essential event in the life of a T cell. Expression of RAG1 and RAG2 is lymphoid-specific and dictates irreversible T cell lineage commitment in developing thymocytes.⁶ Moreover, gene rearrangement provides the T cell compartment with sufficient diversity to sustain protective immunity. Indeed, the importance of receptor diversification is apparent in murine models and in a number of primary immunodeficiencies in humans. For instance, *rag*-deficient mice are devoid of T and B lymphocytes.⁷ In humans, *rag* deficiency is linked to severe combined immunodeficiency, and other *rag* mutations can lead to immunodeficiency with expansion of $\gamma\delta$ T cells or with idiopathic CD4⁺ T cell lymphopenia. Mis-sense mutations in *rag1* and *rag2* are the cause of Omenn syndrome, a disease with graft-versus-host disease-like

¹Cardiff University School of Medicine, Cardiff, UK and ²Department of Pathology, University of Massachusetts Medical School, Worcester, MA, USA
Correspondence: Professor AK Sewell, Division of Infection and Immunity Cardiff University School of Medicine, Henry Wellcome Building, University Hospital Wales, Cardiff, CF14 4XN, UK.

E-mail: sewellak@cardiff.ac.uk

Received: 16 December 2014; Accepted: 17 December 2014

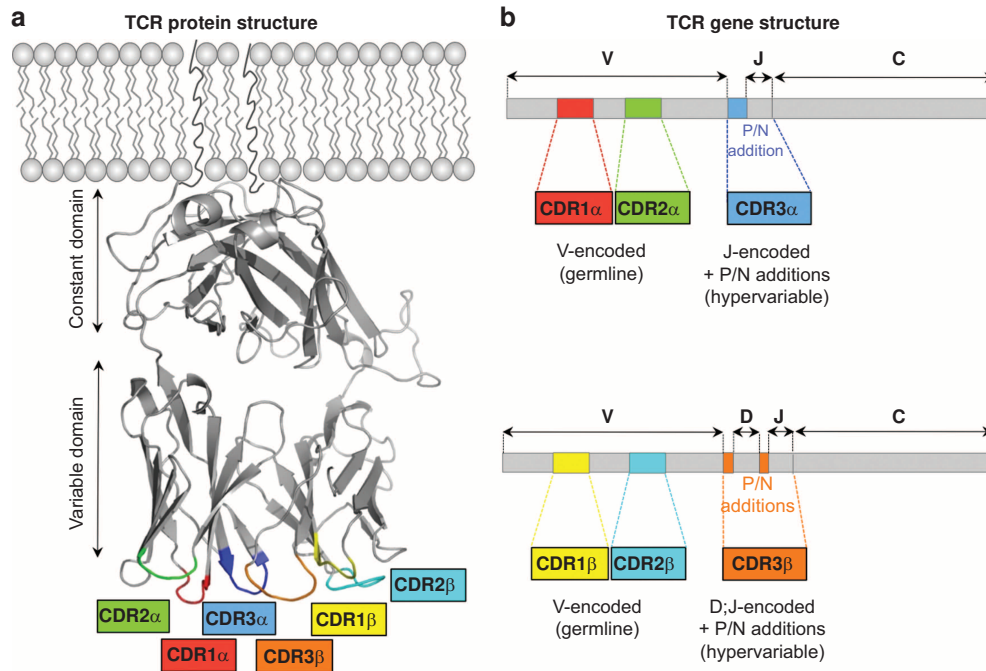


Figure 1 TCR protein and gene structure. (a) Structure of the AS01 TCR.³⁹ The TCR- α and TCR- β CDR loops are short hairpin turns linking adjacent β -strands. (b) CDR1 and CDR2 are entirely encoded in the germline V genes, whereas CDR3 lies at the junction between the rearranged V and J segments (TCR- α) and V, D and J segments (TCR- β). The CDR3 junctional sites are assembled by random addition and deletion of template and non-template nucleotides (blue for TCR- α and orange for TCR- β). CDR, complementarity-determining region; TCR, T cell receptor.

clinical presentation.^{8,9} Omenn syndrome is characterized by expansion of autoreactive CD4⁺ T cells with an oligoclonal repertoire and is fatal to infants between 2 and 6 months of age as a result of recurrent infections.¹⁰

The theoretical diversity in the TCR- $\alpha\beta$ repertoire is estimated at 10^{15} clones in mice¹¹ and 10^{18} in humans.¹² Most of these specificities will never be used during an individual's lifetime, as the murine and human peripheral repertoires are composed of 2 million¹³ and 25 million¹⁴ clonotypes, respectively. Out of 25 million human TCRs, some clonotypes referred to as 'public' TCRs can be found in different individuals, while others are largely unique to an individual and are said to be 'private'.

The molecular principles that dictate which TCRs are assembled and selected to seed the peripheral pool, which are shared between individuals and which are private are only starting to emerge. Dissecting the distribution of TCR clonotypes within an individual, and across individuals, in health and disease is critical to our understanding of protective T cell-mediated responses. In this review, we discuss the various factors working to shape the pre-selection, naive and antigen-specific TCR repertoires. We bring particular attention to recent studies which suggest that TCR 'signatures' shared across genetically disparate individuals may become important diagnostic tools and predictors of immune protection or disease.

TCR DIVERSITY IN THE PRE-SELECTION REPERTOIRE

Gene rearrangement is typically thought of as an inherently random process. Intuitively, stochastic diversification of the repertoire would seem advantageous, maximizing potential immune coverage without prior bias towards certain specificities. However, numerous studies have demonstrated that the complexity of the repertoire is not achieved at random. Rather, generation of diversity in $\alpha\beta$ T cells is tightly regulated and the composition of the repertoire, even prior to thymic selection is highly structured.

Both genetic and epigenetic factors influence the composition of the pre-selection repertoire. The 'accessibility hypothesis'¹⁵ posits that in order for recombination to take place, gene segments must first be made accessible to the recombination machinery. This in turn depends on subnuclear relocation of the rearranging TCR loci (*tr*), DNA methylation status, recruitment of chromatin remodelling enzymes, histone modification and germline transcription. The mechanisms involved in spatial and temporal control of V (D) J recombination have been reviewed elsewhere.¹⁶

Activation of the 3' proximal region of antigen receptor loci is well characterized and known to be dependent on activation of a local enhancer. However, the factors that govern the accessibility and activation of the 5' V region are less clear.^{17,18} In particular, whether differential accessibility and activation status of V genes can affect the composition of the resulting

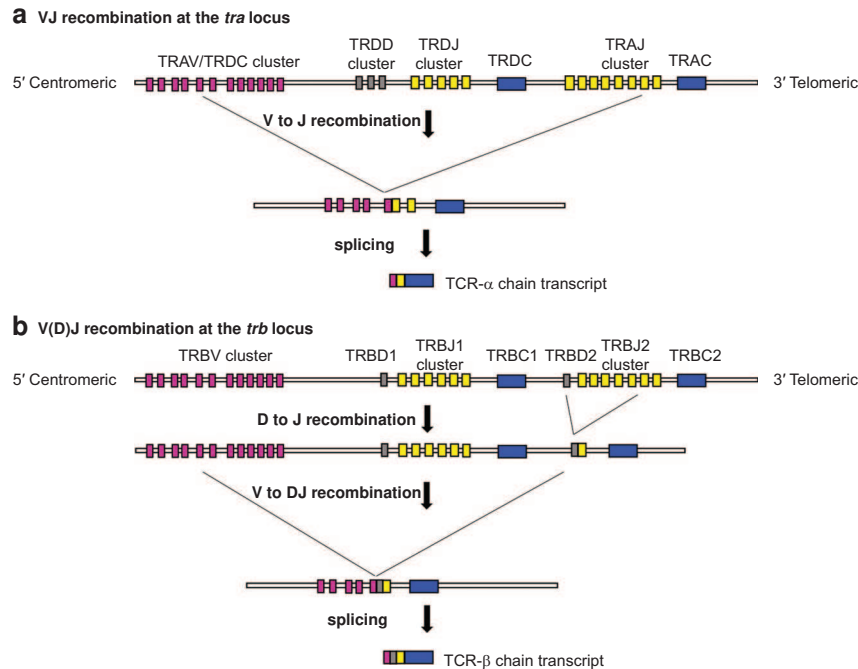


Figure 2 Gene rearrangement at the *tr* loci. **(a)** VJ recombination at the *tra* locus. The *tra* locus (in which the TCR- δ , or *trd* locus is also embedded) comprises a 5' V gene segment cluster (46 TRAV segments) followed by a central J cluster (51 TRAJ segments) and a single C gene segment (TRAC). TCR- δ D and J segments (TRDD and TRDJ, respectively) are also present in the locus. V to J recombination brings together one of many TRAV segments to one of many TRAJ segments. The intervening sequences are spliced out, producing a TCR- α transcript in which V, J and C segments are directly adjacent. **(b)** VDJ recombination at the *trb* locus. The *trb* locus is composed of a 5' V cluster (48 TRBV gene segments) followed by two 3' TRBD–TRBJ–TRBC clusters. VDJ recombination is a two-step, ordered process. D to J recombination occurs first, juxtaposing TRBD1 to one of the six TRBJ1 segments or TRBD2 to one of the seven TRBJ2 segments. V to DJ recombination subsequently brings the rearranged DJ join to one of many TRBV segments. The intervening sequences are then spliced out, generating a TCR- β transcript in which, V, D, J and C segments are directly adjacent. TCR, T cell receptor.

repertoire is largely unknown. In the immunoglobulin heavy chain (*Igh*) locus, distal V genes have been shown to have higher levels of active histone markers compared to proximal segments, which suggests that different V elements recombine at different frequencies despite being equally accessible to the recombination machinery.¹⁹ Whether this is also true for *tr* loci has yet to be determined.

Recently, a comprehensive analysis of the mouse TCR- α repertoire revealed that the frequency of out-of-frame sequences was dependent on V and J segment usage, suggesting that the production of out-of-frame, non-functional TCR- α chains is genetically determined.²⁰ Moreover, syngeneic mice have been shown to display at least 15% overlap in the TCR- β chain repertoire. Most of this overlap is attributed to recombinatorial bias, as the vast majority of shared sequences are observed even in the pre-selection repertoire.²¹ An earlier study looking into the TCR- β chain repertoire of TCR- $\alpha^{-/-}$ mice had already suggested recombinatorial bias in the pre-selection repertoire. In TCR- $\alpha^{-/-}$ mice, the TCR- $\alpha\beta$ receptor is not expressed, providing insight into the composition of the rearranged TCR- β pool in the absence of thymic selection. In this setting, 11% of all analysed sequences were repeats.²²

Altogether, these studies, and others, suggest bias in the composition of the TCR repertoire prior to any contribution

from thymic selection. Notably, V (D) J recombination determines the extent of TCR sharing between different individuals, irrespective of MHC haplotype.

TCR DIVERSITY IN THE NAIVE POOL

Successfully rearranged TCRs are expressed at the T cell surface and audition for selection on thymic self-pMHC ligands. The net result of thymic selection is that the post-selection repertoire is largely purged of most clonotypes.²³ Typically, only one in a hundred thymocytes are thought to be granted access to the periphery²⁴ (Figure 3). Assessing the relative distribution of TCR clonotypes has long been a challenge in the naive pool because of low precursor frequency. Nevertheless, identifying the factors that shape the composition of the naive repertoire is critical to our understanding of protective T cell-mediated immunity because naive lymphocytes represent the precursor pool from which all immune responses arise.

The size and diversity of the post-selection thymocyte population is regulated by the ligands made available in the thymus by antigen processing and presentation.^{25,26} Polymorphism at the *mhc* will affect an individual's TCR repertoire by determining the collection of peptides that can be presented to T cells during development.^{27–29} The role of the MHC-bound peptide in positive selection was made clear from experiments with

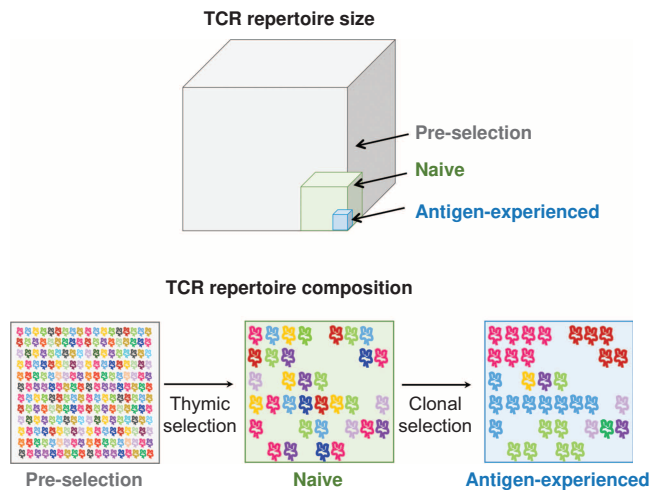


Figure 3 Size and composition of the pre-selection, naive and antigen-experienced repertoires. TCR diversity is greatest in the pre-selection repertoire (gray). Positive and negative selection in the thymus purges the pre-selection repertoire of most specificities, creating a peripheral naive repertoire that is substantially less diverse (green). In the periphery, antigen exposure further narrows the repertoire over time leading to clonal expansion of antigen-specific populations (blue). TCR diversity is largely preserved throughout the human lifespan, except in infancy and old age, but the net distribution of TCR clonotypes is altered. TCR, T cell receptor.

foetal thymic organ cultures in which the major route for class I MHC loading and surface expression was blocked by deficiency in the transporter associated with antigen processing 1. In transporter associated with antigen processing 1-deficient foetal thymic organ cultures, the addition of exogenous peptide enhanced class I expression and improved the selection of mature CD8⁺ T cells.³⁰ Later experiments demonstrated that larger peptide mixtures were better at positively selecting T cells with a broad repertoire than single peptides,³¹ suggesting that peptide diversity in the thymic stroma correlates with TCR diversity.

At least one other component of the antigen processing machinery, ER-associated aminopeptidase 1 (ERAP1) is also highly polymorphic. Single nucleotide polymorphisms in ERAP1 are inherited as haplotypes, some of which are linked to autoimmune disease.³² ERAP1 trims peptides entering the ER in order to increase the frequency of peptides of appropriate length for binding to class I MHC molecules. Of note, ERAP1 activity can lead to peptide over-trimming and destruction. Therefore, allelic variants of ERAP1, with distinct trimming properties, will also influence the composition of the repertoire by determining the identity and abundance of peptides presented to CD8⁺ T cells.³²

Nitta *et al.*³³ have shown that mice deficient in the thymoproteasome subunit $\beta 5t$ select fewer CD8⁺ T cells than $\beta 5t$ -sufficient mice. In this setting, CD8⁺ T cells display reduced TCR diversity, which demonstrates that the peptides generated by the unique catalytic activity of the thymoproteasome affect the composition of the repertoire. Similarly, in knock-in mice

in which the immunoproteasome $\beta 5i$ subunit replaces $\beta 5t$, positive selection is compromised.³⁴

In summary, TCRs auditioning for selection are picked in a deterministic manner. The factors controlling selection encompass some of the major components of the antigen processing machinery, which determine the peptide universe generated in the thymus.

TCR BIAS IN ANTIGEN-SPECIFIC RESPONSES

The thymus involutes with age and thymic output is reduced as a consequence but the composition of the repertoire remains remarkably constant throughout life, except in infancy and old age.³⁵ Homeostatic regulation ensures that most specificities generated by the thymus are maintained during the lifetime of the individual.^{36,37} However, the relative abundance of each specificity is modulated by the individual's history of antigen exposure, as antigen-driven selection in the periphery leads to differential expansion of specific TCR clonotypes. It follows that TCR diversity is the highest in the naive compartment, with the antigen-experienced repertoire being skewed towards just some of these specificities^{36,38} (Figure 3).

In mice, the emergence of a diverse repertoire is a predictor of good disease outcome. Messaoudi *et al.*²⁹ showed that wild-type C57BL/6 mice (H-2^b) infected with *herpes simplex virus-1* select for a significantly narrower repertoire, compared to a co-isogenic strain (H-2^{bm8}) which differed only by expression of a H-2K molecule with four amino acid mutations in the peptide-binding groove. Strikingly, H-2^{bm8} mice showed resistance to infection and increased survival compared to their wild-type counterpart, which led the authors to conclude that the expansion of additional TCRs was necessary and sufficient to confer protection against *herpes simplex virus-1*. Indeed, in this system, *herpes simplex virus-1* resistance was strictly conveyed by V β 8-expressing T cells.

Some human infections paint a different picture. Infections with common pathogens give rise to highly skewed and predictable repertoires with clonotypes reported to be shared across HLA-matched individuals. Thus, the advantage of having a diverse repertoire is not always apparent in human disease. The occurrence of shared TCRs is explained by convergent recombination, whereby certain TCR sequences are produced at high frequency (this is the case for near germline sequences requiring few nucleotide additions), and by the selection of TCRs with a selective advantage such as structural features that are optimal for pMHC recognition.³⁹

In SIV-infected rhesus macaques, the emergence of public Gag-specific clonotypes correlates with protection and the absolute number of shared clonotypes inversely correlates with viral load.⁴⁰ Similarly in humans, control of HIV-1 replication in the absence of antiretroviral therapy is mediated by a few TCR clonotypes which are selected in the context of the so-called 'protective MHC alleles' HLA-B*57, HLA-B*27 and HLA-B*58.⁴¹ These findings imply that certain clonotypes shared across several distinct individuals can be reliably associated with protective, beneficial immune responses and may act as surrogate predictors of disease outcome.

Many T cell malignancies are characterized by extreme clonal skewing, with oligoclonal or even monoclonal T cell expansions. Recently, Clemente *et al.*⁴² showed that in HLA-A2⁺ patients with T-large granular lymphocyte leukaemia, V β 17-expressing T cells share a unique CDR3 sequence. Importantly, sequencing depth in this study was sufficient to establish that this canonical TCR- β sequence associated with T-large granular lymphocyte leukaemia was largely undetectable in healthy individuals. Similarly, another study following the outcome of autologous stem cell transplantation for the treatment of juvenile idiopathic arthritis suggested that TCR- β oligoclonality was linked to clinical relapse.⁴³ Two patients in complete remission displayed diverse TCR- β CDR3 immediately following transplant and at later follow-up time points, whereas the third patient presented with oligoclonal CD8⁺ T cell skewing in most V β families and relapsed within a month. Moreover, in the latter case, the dominant clones were shown to have emerged both from the pre-transplant pool and from *de novo* TCR rearrangement. Thus, because TCR skewing arises in a predictable fashion following an antigen-specific response, monitoring TCR oligoclonality and tracking specific TCR clonotypes linked to malignancy, or other immunological disorders, may prove beneficial in the clinical setting.

TCR DIVERSITY IN HEALTH AND DISEASE

Inbred mice represent a powerful tool for the analysis of pre-selection, immunologically naive and antigen-specific repertoires. Murine models first suggested that in a normal setting, the formation of the TCR repertoire was H-2-dependent⁴⁴ and that TCR usage was altered in the context of autoimmune disease.^{45,46} For instance, the non-obese diabetic mouse regulatory T cell (T_{reg}) repertoire is significantly restricted compared to conventional T cells and to T_{reg} from wild-type mice. Strikingly, this defect in generation of diversity is apparent in *rag*^{-/-} B6 mice reconstituted with non-obese diabetic bone marrow indicating a cell-intrinsic origin, i.e., independent of the nature of the selecting thymic stroma. Furthermore, using different congenic non-obese diabetic strains, it was shown that T_{reg} diversity was in fact regulated by a yet-unidentified gene on chromosome 4.⁴⁷ This suggests that unknown genetic mechanisms may be at work to modulate diversity in the T_{reg} lineage, which in turn is critical to the establishment of tolerance. Similarly in humans, TCR diversity is known to be critical for homeostasis of T_{reg} cells and suppressor function.⁴⁸ Therefore, understanding the mechanisms underlying the formation of the T_{reg} repertoire will further our understanding of autoimmune disease and provide insight into how selection of a dysfunctional repertoire leads to disease.

In humans, twin studies are invaluable tools for dissecting the genetic factors underlying the shaping of the peripheral repertoire. Examples of repertoire analysis in twins, particularly hyperanalytical methods such as high-throughput sequencing, are still scarce in the literature. Nonetheless, a few select reports are already providing key information. The earliest study of TCR- β usage by Gulwani-Akolkar *et al.*⁴⁹ already highlighted the influence of HLA alleles. HLA-identical siblings

were found to have the highest degree of similarity in the TCR- β repertoire, whereas HLA-haploidentical or HLA-mismatched siblings were dissimilar. Later, Davey *et al.*⁵⁰ re-examined this finding in seven pairs of monozygotic twins. The highest degree of similarity was found in the youngest pair (aged 2) and all other twins (aged 5–44) had at least one difference in V β segment usage. Four pairs of twins in this study had a history of disease affecting either one (discordant) or both individuals (concordant). In discordant pairs, notable differences were seen in CD8⁺ T cell V β usage. For instance, one individual suffering from asthma showed a significant increase in V β 8 usage and loss of V β 12, compared to their healthy twin. Differences in V β expression patterns were also noted in two other pairs of discordant twins affected by Hodgkin's lymphoma and by polycythemia vera. Interestingly, V β usage was also found to diverge in twins concordant for systemic lupus erythematosus. The authors therefore suggested that both environmental factors and genetic factors contributed substantially to shaping the TCR repertoire. Thus, at birth, monozygotic twins have near identical repertoires with respect to V β usage, but changes associated with an individual's unique history of antigen exposure arise over time.

A longitudinal study following monozygotic twins simultaneously infected with the same HIV-1 strain by intravenous drug abuse suggested random TCR recruitment in various epitope-specific responses. Over the course of the disease and virus evolution, several Pol and Nef epitopes were found to be concordant (shared between the twins), whereas others were unique to either individual. Remarkably, in these individuals, the TCR- β chain repertoires raised against the shared epitopes were exclusively private. Thus, the identical genetic background and the concordant pathways of viral escape did not lead to TCR sharing in this instance and the authors concluded that the recruitment of pathogen-specific TCRs was essentially stochastic.⁵¹ This finding therefore suggested that an individual's natural history of antigen exposure may be at least as important as their genetic background in determining the composition of the antigen-specific repertoire and that the forces shaping the repertoire are more complex than previously thought.

TCR CLONOTYPES AS MARKERS OF DISEASE

Tracking and detecting TCR 'signatures' associated with specific human infections may prove problematic considering the complexity and diversity of human pathogens. Even in cases where TCR expansions can be reliably and accurately linked to infection, developing diagnostic tools based on TCR clonotyping may be highly impractical due to rapid onset and progression of disease, particularly in the case of acute infections. Nonetheless, this approach may become an option in the context of autoimmune disease or cancer. Indeed, for such disorders, pathogenesis is usually slow and, at least in the case of autoimmunity, the onset of symptoms and formal diagnosis can take place several years after autoantibodies are first detectable in serum. TCR-based diagnoses might offer several practical advantages, as technical resolution is rapidly increasing

with the development of high-throughput 'next generation' sequencing methodologies.

As discussed above, a number of studies suggest that V β usage is fairly predictable, based on the influence of HLA alleles. Nevertheless, understanding how the composition of the TCR repertoire is altered in the disease setting remains a challenging task due to technical limitations on the one hand, but also to the complex functional and genetic nature of the T cell pool on the other hand. Studies of monozygotic twins that are discordant for autoimmune diseases have been particularly enlightening.

Fozza *et al.*⁵² looked at the repertoires of monozygotic twins discordant for type 1 diabetes at the time of the study. In this study, any individual was found to share more TCR- β clonotypes with their twin than with unrelated individuals. However, the extent to which TCR- β repertoires overlapped was unlinked to disease status. This is to be expected because the presence of a given TCR sequence does necessarily implicate a particular T cell clone in disease pathogenesis. Indeed, using CD8⁺ T cell clones derived from a pair of multiple sclerosis (MS)-discordant twins, Somma *et al.*⁵³ identified a number of TCR- β implicated in the pathogenesis of MS which could be clonally expanded either from the healthy or the affected twin.

In contrast, Utz *et al.*⁵⁴ showed that bulk T cell cultures harvested from monozygotic twins concordant for MS mobilized the same TCR- α chains following challenge with myelin basic protein, whereas TCR- α chains from discordant twins were dissimilar. Another study has recently demonstrated that the cerebrospinal fluid of MS patients is enriched with public EBV-cross-reactive TCR clonotypes which, again, could serve as potential markers of disease.⁵⁵ Thus, in these examples, similarities in V α and V β usage were largely linked to genetic factors and not critically influenced by nominal antigen. Differential TCR usage in discordant twins could mirror the role of T cell-mediated responses as drivers of autoimmune disease and mapping how the TCR repertoire is altered by various disease settings will be key to the establishment of TCR-based diagnostic methods. The discrepancies in the aforementioned studies highlight the need for more rigorous and more powerful sequencing approaches coupled to the identification of antigen-experienced T cell subpools. Indeed, developments in T cell phenotyping by flow cytometry are also moving apace, making it possible to identify specific TCRs within a given T cell subpopulation. Determining the frequency and distribution of particular TCR sequences among discrete T cell subcompartments might yet prove to be extremely powerful.⁵⁶

Using 'next-generation' sequencing, a more recent study in three pairs of healthy, monozygotic twins suggested that the composition of the TCR repertoire at the CDR3 level was also largely determined by genetic factors.⁵⁷ V segment usage in out-of-frame sequences was found to be more similar in twins than in non-twins, indicating a genetically determined bias prior to selection, in line with had been previously reported by Genolet *et al.*²⁰ Remarkably, monozygotic twins did not

share more in-frame sequences with their twin than with an unrelated individual, indicating that certain TCR clonotypes can be enriched and shared across a wide population irrespective of HLA type, as previously suggested by Robins *et al.*³⁸ This is in accordance with previous studies highlighting convergent recombination as a major route for the generation of public TCRs. Some of the sequences shared by multiple, unrelated individuals may be enriched with clonotypes derived from invariant populations such as invariant natural killer T cells, mucosa-associated invariant T cells and germline-encoded, mycoyl-reactive T cells.⁵⁸ Such semi-invariant TCRs are likely raised against common antigens and shared across the population. Thus, if linked to certain infections, such TCRs could become invaluable tools for the diagnosis of human disease (Figure 4).

Some diseases are associated with the emergence of aberrant clonotypes. All *tr* loci normally rearrange in *cis*, or in other words, strictly within locus. In pathologies linked to chromosomal instability including various malignancies, *trans* rearrangements arise in the periphery at high frequency. Inversions on chromosome 7 can give rise to V γ to D β J β recombination. Such rearrangements represent less than 1 in 100 000 peripheral blood lymphocytes in healthy individuals. However, in patients with ataxia telangiectasia, the abundance of such clonotypes is increased 50- to 100-fold. Similarly, patients with childhood acute lymphoid leukaemia of B-cell lineage present with high frequency of IGHV/J α hybrids.⁵⁹ In various lymphoma patients, the V(D)J recombination machinery has been implicated in abnormal interchromosomal joining. Thus, *trans* rearrangements are abundant in such patients but either largely absent or extremely rare in healthy individuals and may become important diagnostic markers for several disorders and malignancies (Figure 4).

CONCLUSIONS

T cells orchestrate a variety of immune responses against self- and foreign antigen and the genetic basis of these responses is encrypted in the TCR repertoire. Over the last two decades, numerous studies have accumulated in the literature to suggest that the composition of the TCR repertoire is a tightly regulated quantity. The formation of the pre-selection and the naive repertoires is largely determined by genetic factors, including HLA type and genes encoding the components of the V(D)J recombination and antigen processing machineries. Thus, against a given haplotype, the TCR repertoire is a structured immunological system. However, the statistical distribution of TCR clonotypes incurs large changes over time according to the personal history of antigen encounter. In some cases, even HLA-identical monozygotic twins can raise different TCR pools against the same antigen. Consequently, the TCR repertoire, albeit regulated, and governed by genetic forces, is nonetheless a large and complex system. How the repertoire differs between distinct individuals, whether genetically identical or genetically disparate, is poorly understood and dissecting the composition of the human repertoire in the antigen-experienced compartment has long remained a challenge due to technical

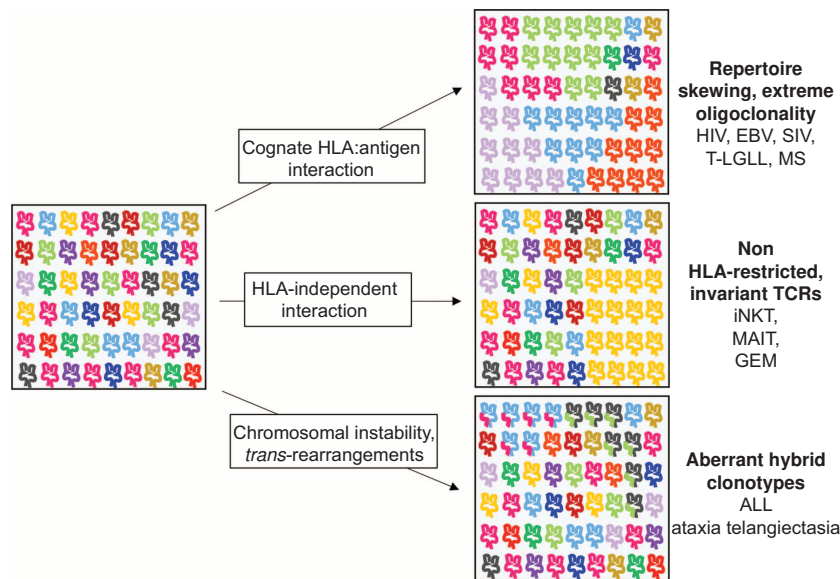


Figure 4 Skewing of the TCR repertoire in human disease. The peripheral TCR repertoire is shaped by antigen encounter and altered in the context of disease. Classical pMHC recognition leads to clonal expansion of antigen-specific T cells, which in some human pathologies can lead to extreme oligoclonality and skewing (top). In this setting, the expansion of public clonotypes can be beneficial as seen in HIV-1 infection, but in other cases, certain clonotypes are involved in disease pathogenesis as described for MS. Other semi-invariant clonotypes such as iNKT (also called NKT type I), MAIT and GEM TCRs expand in response to some microbial infections in an HLA-independent manner (centre). Some malignancies such as ALL, or other disorders associated with chromosomal instability, provoke the expansion of aberrant clonotypes (Ig/TCR hybrids or TCR- γ /TCR- β hybrids; bottom) that are largely absent from the healthy. ALL, acute lymphoid leukaemia; GEM, germline-encoded, mycoyl-reactive; iNKT, invariant natural killer T; MAIT, mucosa-associated invariant T; MS, multiple sclerosis; pMHC, peptide-major histocompatibility complex; TCR, T cell receptor.

limitations. The methodology and the tools required for comprehensive genetic analysis of antigen receptors have only recently begun to unravel the complexity of the T cell compartment. Notably, numerous studies have described how the repertoire might be altered in the context of certain infections, malignancies or immunological disorders. One trending theme is that some pathologies provoke the emergence of specific TCR clonotypes, which may prove to be invaluable immunological 'signatures' in the clinical setting. Such signatures include public TCR chains raised against common pathogens, extreme oligoclonality and TCR skewing as seen in many lymphoid malignancies, unique clonotypes associated with autoimmune disease or even aberrant TCR hybrid chains that are linked to numerous immunological disorders. These may represent only a few examples of the information encoded in the TCR repertoire and our understanding of the forces governing the complexity of the T cell compartment in health and disease may be key to future diagnostic and therapeutic interventions.

ACKNOWLEDGEMENTS

Andrew Sewell and Meriem Attaf are supported by a grant from the Wellcome Trust (WT100327MA).

- 1 Alam SM, Davies GM, Lin CM, Zal T, Nasholds W, Jameson SC *et al*. Qualitative and quantitative differences in T cell receptor binding of agonist and antagonist ligands. *Immunity* 1999; **10**: 227–237.

- 2 Almeida A, Rocha B, Freitas AA, Tanchot C. Homeostasis of T cell numbers: from thymus production to peripheral compartmentalization and the indexation of regulatory T cells. *Sem Immunol* 2005; **17**: 239–249.
- 3 Kappler J, Kubo R, Haskins K, Hannum C, Marrack P, Pigeon M *et al*. The major histocompatibility complex-restricted antigen receptor on T cells in mouse and man: identification of constant and variable peptides. *Cell* 1983; **35**: 295–302.
- 4 Raullet DH. The structure, function and molecular genetics of the $\gamma\delta$ T cell receptor. *Annu Rev Immunol* 1989; **7**: 175–207.
- 5 Chothia C, Boswell DR, Lesk AM. The outline structure of the T cell $\alpha\beta$ receptor. *EMBO J* 1988; **7**: 3745.
- 6 Yui MA, Rothenberg EV. Developmental gene networks: a triathlon on the course to T cell identity. *Nat Rev Immunol* 2014; **14**: 529–545.
- 7 Shinkai Y, Rathbun G, Lam KP, Oltz EM, Stewart V, Mendelsohn M *et al*. RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* 1992; **68**: 855–867.
- 8 Corneo B, Moshous D, G ng r T, Wulffraat N, Philippet P, Le Deist FL *et al*. Identical mutations in RAG1 or RAG2 genes leading to defective V(D)J recombination activity can cause either T-B-severe combined immune deficiency or Omenn syndrome. *Blood* 2001; **97**: 2772–2776.
- 9 Villa A, Sobacchi C, Notarangelo LD, Bozzi F, Abinun M, Abrahamsen TG *et al*. V(D)J recombination defects in lymphocytes due to RAG mutations: severe immunodeficiency with a spectrum of clinical presentations. *Blood* 2001; **97**: 81–88.
- 10 Aleman K, Noordzij JG, de Groot R, van Drogen JJ, Hartwig NG. Reviewing Omenn syndrome. *Eur J Pediatr* 2001; **160**: 718–725.
- 11 Davis MM, Bjorkman PJ. T cell antigen receptor genes and T cell recognition. *Nature* 1988; **334**: 395–402.
- 12 Sewell AK. Why must T cells be crossreactive? *Nat Rev Immunol* 2012; **12**: 669–677.
- 13 Casrouge A, Beaudoin E, Dalle S, Pannetier C, Kanellopoulos J, Kourilsky P. Size estimate of the alpha beta TCR repertoire of naive mouse splenocytes. *J Immunol* 2000; **164**: 5782–5787.

- 14 Arstila T, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P. A direct estimate of the human $\alpha\beta$ T cell receptor diversity. *Science* 1999; **286**: 958–961.
- 15 Yancopoulos GD, Alt FW. Developmentally controlled and tissue-specific expression of unrearranged VH gene segments. *Cell* 1985; **40**: 271–281.
- 16 Ebert A, Medvedovic J, Tagoh H, Schwickert TA, Busslinger M. Control of antigen receptor diversity through spatial regulation of V(D)J recombination. *Cold Spring Harb Symp Quant Biol* 2013; **78**: 11–21.
- 17 Krangel MS, McMurry MT, Hernandez-Munain C, Zhong XP, Carabana J. Accessibility control of T cell receptor gene rearrangement in developing thymocytes. The TCR alpha/delta locus. *Immunol Res* 2000; **22**: 127–135.
- 18 Mathieu N, Hempel WM, Spicuglia S, Verthuy C, Ferrier P. Chromatin remodelling by the T cell receptors (TCR)- β gene enhancer during early T cell development: implications for the control of TCR- β locus recombination. *J Exp Med* 2000; **192**: 625–636.
- 19 Choi NM, Loguercio S, Verma-Gaur J, Degner SC, Torkamani A, Su AI *et al*. Deep sequencing of the murine IgH repertoire reveals complex regulation of nonrandom V gene rearrangement frequencies. *J Immunol* 2013; **191**: 2393–2402.
- 20 Genolet R, Stevenson BJ, Farinelli, Østerås M, Luescher IF. Highly diverse TCR α chain repertoire of pre-immune CD8⁺ T cells reveals new insights in gene recombination. *EMBO J* 2012; **31**: 1666–1678.
- 21 Li H, Ye C, Ji G, Wu X, Xiang Z, Li Y *et al*. Recombinatorial biases and convergent recombination determine interindividual TCR β sharing in murine thymocytes. *J Immunol* 2012; **189**: 2404–2413.
- 22 Furmanski AL, Ferreira C, Bartok I, Dimakou S, Rice J, Stevenson FK *et al*. Public T cell receptor beta-chains are not advantaged during positive selection. *J Immunol* 2008; **180**: 1029–1039.
- 23 van Laethem F, Tikhonova AN, Singer A. MHC restriction is imposed on a diverse T cell receptor repertoire by CD4 and CD8 co-receptors during thymic selection. *Trends Immunol* 2012; **33**: 437–441.
- 24 Goldrath AW, Bevan MJ. Selecting and maintaining a diverse T cell repertoire. *Nature* 1999; **402**: 255–262.
- 25 Nikolich-Zugich J, Bevan MJ. Role of self-peptides in positively selecting the T cell repertoire. *Nature* 1990; **344**: 65.
- 26 Anderson G, Partington KM, Jenkinson EJ. Differential effects of peptide diversity and stromal cell type in positive and negative selection in the thymus. *J Immunol* 1998; **161**: 6599–6603.
- 27 Bevan MJ. In thymic selection, peptide diversity gives and takes away. *Immunity* 1997; **7**: 175–178.
- 28 Dyal R, Messaoudi I, Janetzki S, Nikolich-Zugich J. MHC polymorphism can enrich the T cell repertoire of the species by shifts in intrathymic selection. *J Immunol* 2000; **164**: 1695–1698.
- 29 Messaoudi I, Guevara Patino A, Dyal R, LeMaout J, Nikolich-Zugich J. Direct link between *mhc* polymorphism, T cell avidity, and diversity in immune defense. *Science* 2002; **298**: 1797–1800.
- 30 Ashton-Rickardt P, van Kaer L, Schumacher TN, Ploegh HL, Tonegawa S. Peptide contributes to the specificity of positive selection of CD8⁺ T cells in the thymus. *Cell* 1993; **73**: 1041.
- 31 Hogquist KA, Gavin MA, Bevan MJ. Positive selection of CD8⁺ T cells induced by major histocompatibility complex binding peptides in fetal thymic organ culture. *J Exp Med* 1994; **177**: 1469–1473.
- 32 Reeves E, Edwards CJ, Elliott T, James E. Naturally occurring *ERAP1* haplotypes encode functionally distinct alleles with fine substrate specificity. *J Immunol* 2011; **191**: 35–43.
- 33 Nitta T, Murata S, Sasaki K, Fujii H, Ripen A, Ishimaru N *et al*. Thymoproteasome shapes immunocompetent repertoire of CD8⁺ T cells. *Immunity* 2010; **32**: 29–40.
- 34 Xing Y, Jameson SC, Hogquist KA. Thymoproteasome subunit- β 5T generates peptide-MHC complexes specialized for positive selection. *Proc Natl Acad Sci USA* 2013; **110**: 6979–6984.
- 35 Blackman M, Woodland DL. The narrowing of the CD8 T cell repertoire in old age. *Curr Opin Immunol* 2011; **23**: 537–542.
- 36 Rudd BD, Venturi V, Davenport MP, Nikolich-Zugich J. Evolution of the antigen-specific CD8⁺ TCR repertoire across the life span: evidence of clonal homogenization of the old TCR repertoire. *J Immunol* 2011; **186**: 2056–2064.
- 37 Qi Q, Liu Y, Cheng Y, Glanville J, Zhang D, Lee JY *et al*. Diversity and clonal selection in the human T cell repertoire. *Proc Natl Acad Sci USA* 2014; **111**: 13139–13144.
- 38 Robins HS, Srivastava SK, Campregher PV, Turtle CJ, Andriesen J, Riddell SR *et al*. Overlap and effective size of the human CD8⁺ T cell receptor repertoire. *Sci Transl Med* 2010; **2**: 47ra64.
- 39 Miles JJ, Bulek AM, Cole DK, Gostick E, Schauenburg AJ, Dolton G *et al*. Genetic and structural basis for selection of a ubiquitous T cell receptor deployed in Epstein–Barr virus infection. *Plos Pathog* 2010; **6**: e1001198.
- 40 Price DA, Asher TE, Wilson NA, Nason MC, Brenchley JM, Metzler IS *et al*. Public clonotype usage identifies protective Gag-specific CD8⁺ T cell responses in SIV infection. *J Exp Med* 2009; **206**: 923–936.
- 41 Chen H, Ndhlovu ZM, Liu D, Porter LC, Fang JW, Darko S *et al*. TCR clonotypes modulate the protective effect of HLA class I molecules in HIV-1 infection. *Nat Immunol* 2012; **13**: 691–702.
- 42 Clemente MJ, Przychodzen B, Jerez A, Dienes BE, Aftable MG, Hussein-zadeh H *et al*. Deep sequencing of the T cell receptor repertoire in CD8⁺ T-large granular lymphocyte leukemia identifies signature landscapes. *Blood* 2013; **122**: 4077–4085.
- 43 Wu Q, Pesenacker AM, Stansfield A, King D, Barge D, Poster HE *et al*. Immunological characteristics and T cell receptor clonal diversity in children with systemic juvenile idiopathic arthritis undergoing T cell-depleted autologous stem cell transplantation. *Immunology* **142**: 227–236.
- 44 von Boehmer H, Teh HS, Kisielow P. The thymus selects the useful, neglects the useless and destroys the harmful. *Immunol Today* 1989; **10**: 57–61.
- 45 Maeda T, Sumida T, Kurasawa K, Tomioka H, Itoh I, Yoshida S *et al*. T-lymphocyte-receptor repertoire of infiltrating T lymphocytes into NOD mouse pancreas. *Diabetes* 1991; **40**: 1580–1585.
- 46 Osman GE, Toda M, Kanagawa O, Hood L. Characterization of the T cell receptor repertoire causing collagen arthritis in mice. *J Exp Med* 1993; **172**: 387–395.
- 47 Ferreira C, Palmer D, Blake K, Garden OA, Dyson J. Reduced regulatory T cell diversity in NOD mice is linked to early events in the thymus. *J Immunol* 2014; **192**: 4145–4152.
- 48 Föhse L, Suffner J, Suhre K, Wahl B, Lindner C, Lee CW *et al*. High TCR diversity ensures optimal function and homeostasis of Foxp3⁺ regulatory T cells. *Eur J Immunol* 2011; **41**: 3101–3113.
- 49 Gulwani-Akolkar B, Posnett DN, Janson CH, Grunewald J, Wigzell H, Akolkar P *et al*. T cell receptor V-segment frequencies in peripheral blood T cells correlate with human leukocyte antigen type. *J Exp Med* 1991; **174**: 1139–1146.
- 50 Davey MP, Meyer MM, Bakke AC. T cell receptor V β gene expression in monozygotic twins, discordance in CD8 subset and in disease states. *J Immunol* 1994; **152**: 315–321.
- 51 Yu XG, Lichterfeld M, Williams KL, Martinez-Picado J, Walker BD. Random T cell receptor recruitment in human immunodeficiency virus type 1 (HIV-1)-specific CD8⁺ T cells from genetically identical twins infected with the same HIV-1 strain. *J Virol* 2007; **81**: 12666–12669.
- 52 Fozza C, Contini S, Corda G, Viridis P, Galleu A, Bonfigli S *et al*. T cell receptor repertoire analysis in monozygotic twins concordant and discordant for type 1 diabetes. *Immunobiology* 2012; **217**: 920–925.
- 53 Somma P, Ristori G, Battistini L, Cannoni S, Borsellino G, Diamantini A *et al*. Characterization of CD8⁺ T cell repertoire in identical twins discordant and concordant for multiple sclerosis. *J Leukoc Biol* 2007; **81**: 696–710.
- 54 Utz U, Biddison WE, MacFarland HF, McFarlin DE, Flerlage M, Martin R. Skewed T cell receptor repertoire in genetically identical twins correlates with multiple sclerosis. *Nature* 1993; **364**: 243–247.
- 55 Lossius A, Johansen JN, Vardtal F, Robins H, Juratė Saltytė, Holmøy T *et al*. High-throughput sequencing of TCR repertoires in multiple sclerosis reveal intrathecal enrichment of EBV-reactive CD8⁺ T cells. *Eur J Immunol* 2014; **44**: 3439–3452.
- 56 Skowera A, Ladell K, McLaren JE, Dolton G, Matthews KK, Gostick E *et al*. β -cell-specific CD8 T cell phenotype in type 1 diabetes reflects chronic autoantigen exposure. *Diabetes* 2014; in press.
- 57 Zvyagin IV, Pogorelyy MV, Ivanova ME, Komech EA, Shugay M, Bolotin DA *et al*. Distinctive properties of identical twins' TCR

repertoires revealed by high-throughput sequencing. *Proc Natl Acad Sci USA* 2014; **111**: 5980–5985.

- 58 Rossjohn J, Gras S, Miles JJ, Turner SJ, Godfrey DI, McCluskey J. T cell antigen receptor recognition of antigen-presenting molecules. *Annu Rev Immunol* 2014; **33**: in press.
- 59 Allam A, Kabelitz D. TCR trans rearrangements: biological significance in antigen recognition versus the role as lymphoma marker. *J Immunol* 2006; **176**: 5707–5712.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>